DOCUMENT RESUME

ED 329 268 IR 053 459

AUTHOR Hopkinson, Alan
TITLE International Exchange Formats.
PUB DATE 90
NOTE 18p.; Paper presented at the International Symposium on Information Technology (Bangkok, Thailand, September 4-8, 1989).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bibliographic Databases; *Bibliographic Records; International Communication; *Machine Readable Cataloging; *Standards
IDENTIFIERS Common Communication Format; MARC; UNESCO; UNISIST

ABSTRACT

This paper focuses on the establishment of three standard international formats for the exchange of bibliographic data--UNIMARC, CCF, and the UNISIST Reference Manual--and outlines their common and differing features. The development of the UNIMARC manual as the standard international MARC network exchange format is traced, and its salient features, linking techniques, and the UNIMARC companion authorities format are examined. It is noted that criticisms of UNIMARC include subject redundancy in the manual, an incompatibility between record cataloging formats, and its lack of implementation among specific user groups. The establishment of the Common Communication Format (CCF) by Unesco in response to UNIMARC's incompatibilities with other international formats is then documented, and its aim--to establish the exchange of records between both library and secondary service communities--is described. Noting that, in a similar context, automation in the secondary service communities required a standard set of data elements for the exchange of bibliographic data in machine-readable form, this report also describes the development of the UNISIST reference manual and its history and use, as well as CCF users and technical features. A discussion of the CCF in terms of its relationship with existing formats and as an exchange format for bibliographic data concludes the report. (24 references) (MAB)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED329268

# International Exchange Formats

Alan Hopkinson

Paper presented at the
International Symposium on Information Technology
(Bangkok, Thailand September 4-8, 1989)

IR053459

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Alan Hopkinson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

# INTERNATIONAL EXCHANGE FORMATS

A. Hopkinson

## INTRODUCTION

Anyone who is even a little acquainted with standards for the exchange of bibliographic data will know that there are a number of standard formats used for this kind of data transfer. Probably the most used and best known are national MARC formats, USMARC, UK MARC, AUSMARC, MALMARC, etc. In order to exchange data between these, an international MARC format known as UNIMARC has been developed. Other organizations, particularly the secondary services, use the UNISIST Reference Manual. And more recently, the Unesco Common Communication Format has been widely promoted.

This paper concentrates on the international formats and outlines their common and differing features.

## UNIMARC: THE STANDARD INTERNATIONAL MARC NETWORK EXCHANGE FORMAT

UNIMARC was the idea of IFLA. It was conceived of as a tool for an International MARC Network. Although the record structure used by all these formats, which was eventually adopted as international standard ISO 2709 [1], was accepted early on, during the very first co-operative project between the Library of Congress (LC) and the British National Bibliography (BNB, later British Library Bibliographic Services), there had been disagreement on the fields or content designators as they are called between LC and BNB and later between other national libraries. In 1971, a recommendation was made to IFLA that they be responsible for establishing an international standard for content designators. In August 1972, at the IFLA General Conference in Budapest, the IFLA Committee on Cataloguing and the IFLA Committee on Mechanization jointly sponsored the IFLA Working Group on Content Designators. This Working Group had the task of exploring the reasons for the differences between the different MARC formats and arriving at a standard for the international exchange of data in machine-readable form. It limited its investigations to the requirements of the library community, i.e. libraries and national bibliographies. However, to ensure coordination of efforts as widely as possible, all working papers were submitted to the ISO TC46/SC4 Working Group on Content Designators as well as to the UNISIST Working Group on Bibliographic Data Exchange which were both involved with formats for the secondary services. During deliberations, it was realised that each country needed to retain or establish its own format because of differences between national requirements, relating partly to the fact that national bibliographic agencies differed from each other in their roles and partly because of the language barriers that exist between nations. Each national agency would also arrange for the development of conversion programs to convert the data in its own national format into that of the international format. One feature that was agreed on was that the International Standard Bibliographic Descriptions should be the basis of the data

IR053459

elements relating to the descriptive area of the catalogue record. This was a wise move; not only were the ISBDs becoming the basis of national cataloguing codes; their adoption in UNIMARC gave the new format an international flavour and a reference point which librarians not yet familiar with automation could understand. Another feature that was agreed upon was that it should eventually be hospitable to all materials. This was a departure from the Library of Congress practice of having a format for each different type of material and one that gave UNIMARC an advantage over other national formats when a country newly developing a national format sought a model on which to base it. UNIMARC was published in 1977 and the second edition of UNIMARC was published in 1980. This new edition was spurred on by the completion of the ISBDs for cartographic materials, and non-book materials and by the revision of the ISBDs for Monographs and Serials. In the 2nd edition of UNIMARC it states that: "A number of national libraries including those of Australia, Canada, Japan, Hungary, South Africa, the United Kingdom and the United States have already agreed to use UNIMARC as their exchange format with implementation to take place early in the 1980's. To facilitate this the International MARC Network Study, which has already authorized and published several studies relating to the developing network of automated national libraries is giving priority to further studies required to assist the conversion of national MARC formatted data to UNIMARC format." [2] As a token contribution to compatibility at a wider level, Dorothy Anderson, as Director of the IFLA International Office for Universal Bibliographic Control and publisher of UNIMARC responsible for editorial work on the document, persuaded the Working Group to allow her to indicate with an asterisk the data elements regarded as mandatory for the identification and description of a bibliographic item by the Ad hoc Group on the Establishment of the Common Communication Format.

The International MARC Network Study was in the meantime placed under the umbrella of a subgroup of the Conference of Directors of National Libraries called the International MARC Network Study Steering Committee or alternatively the International MARC Network Advisory Committee. The IFLA UBC Office continued to publish papers relating to the study which henceforward were under the authorship of this subgroup. UNIMARC remained an important preoccupation of the group and the format became less the intellectual property of the IFLA Committees on Cataloguing and Mechanization though members of those Committees continued to be involved as members of staff of national libraries interested in UNIMARC.

UNIMARC manual

After the second edition of UNIMARC was published, work began on a UNIMARC interpretive handbook which was later published as the UNIMARC handbook [3]. This uncovered a number of problems in UNIMARC and so a revision was made of the UNIMARC format and of the guidelines and these were published in the UNIMARC manual [4] which became the 3rd edition of the UNIMARC format.

Also, during the 1980s, a review had taken place on the ISBDs for Cartographic Material, Monographic Material, Non-Book Materials and Serials. Described as a "harmonization process", the review was designed to ensure consistency, to provide further and more varied examples, to consider the particular problems of non-

Roman scripts and to modify ISBD(NBM) to make it hospitable to many kinds of material without its assuming the function of a cataloguing code. It was completed in 1986 and though the four ISBDs were not published until 1987 and 1988, they were in a definite enough state to be considered in the revised UNIMARC manual. As this was the third edition of UNIMARC, the format ceased to be contained in a basic standard-like document, but was embedded in its interpretive document. Although, it was expected that this edition will herald a period of relative stability for UNIMARC, nevertheless, some revision will be required in the future. A group is examining the ISBDs for Antiquarian Materials, Printed Music and Computer Files to ensure harmonization. ISBD(G) will be scrutinized to see if any adjustments are needed as a result of the review programme.

## UNIMARC - TECHNICAL DETAILS

UNIMARC was designed on the basis of a set of nine principles which were published in the different editions as 'Guidelines for Format Design'. These were based on experience which had been gained in the different national MARC formats and are too detailed to include here.

### Characteristic features relating to UNIMARC as an exchange format

An interesting features of the format is the inclusion of fields in blocks defined by type of data element. Up to the development of UNIMARC, the major national MARC formats had ordered the different fields in a way that reflected the order of the fields on a traditional catalogue card. UNIMARC avoided this bias towards one particular end product of a machine-readable bibliographic record and put all name access points in one block instead of supplying different fields for author as main entry from author as added entry.

All title access points are defined in the 500 block other than title proper which is field 200 which begins the descriptive block as the title is usually required in the same form as an access point as it is displayed in the descriptive area.

The 100 block is for coded data. Field 100 includes codes common to all materials and each type of material has another field for codes specific to that type.

### Linking techniques

The most novel feature of UNIMARC is its treatment of links between one bibliographic item and another.

Bibliographic items have relationships with each other. They may have previous editions, they may, as in the case of serials, have related, earlier or later titles. Moreover, they may be in the same journal or series as each other. In special cases, some bibliographic items are translations of others.

Another kind of relation is the sharing of common subject or authorship.

UNIMARC has a number of different ways of showing these linking relationships.

Relationships between bibliographic items are indicated by means of fields in the linking entry block, fields 410 to 488. The

largest number of these relate to serials, such as "Continues", "Continues in part", "Changed back to", "Merged with x and y to form". The names of these linking fields are in fact the text that would be associated with the name of the serial in a note generated for the link in a traditional catalogue record.

for serials are "Supplement", "Parent of supplement" and "Issued with".

For monographs and serials are the fields "Series" and "Subseries". These can be used in monographs and serials to link to a containing series and subseries. Links can be made to other editions and to translations or from a translation to its original. These may apply to both monographs and series.

There is additionally a set of linking fields entitled "Levels" which enable links to be made between items in a bibliographic hierarchy. These link to Set, Subset, Piece and Piece-analytic. Since processing of records containing hierarchical links is more complex, character position 8 in the record label is reserved to indicate if this technique has been used. Organizations which had not developed conversion programs for records including these links can thus be warned that they will not be able to process them correctly. Also, it shows that other records will be required for the complete processing of the record that contains these fields. This code has been adopted from character position 19 of the US MARC leader.

In all these cases, the linking fields can be used in two different ways. A link can be made to another record, or the data relating to the related record can be embedded in the linking field. Since one of the main aims of MARC records is to produce catalogue records in printed form, an indicator, the second indicator, specifies whether the field is to be used to print a note: the first indicator is always blank.

Following the indicators, the subfield identifier is $1. There then follows, if a link is being made to a record control number, the record control number preceded for identification by 001, the tag for the record control number or identifier.

If the embedded record technique is used, each field in the embedded record follows the tag which indicates the relation and each field is preceded by $1. These embedded fields are not found in any directory, so processing of these fields in the embedded record is quite different from processing of fields in the main body of the record.

In the record for the serial 'Bus and coach' which was preceded by 'Motor transport' would appear in field 434 the following:

```
_1$15300_$aBus & coach   ['_' represents space]
```

The first two characters are indicators of field 434.
$1 indicates start of the first embedded field 530
0_ are indicators in the embedded field
$aBus & coach are the data which follow immediately.

434 occurs in the directory with pointers to the data string shown above.

If a link were being made to a record number and the record number of 'Bus & coach" was T01564, then the field would appear as follows:

_1$1001T01564

UNIMARC was the first in the family of MARC formats to include this kind of linking mechanism. Hitherto, formats had indicated relationships in other ways, and these methods are retained in UNIMARC itself.

In a traditional catalogue, series relationships are indicated by means of added entries. An item in a monographic series will have an added entry under the name of the series and, if applicable, the number within that series. The series statement which is part of the description of the monograph according to traditional cataloguing practice may be used as an access point if it is the established form. Otherwise, field 410 must be used to contain an embedded record relating to the series. The embedded record may consist of the title of the series; or it may include both author and title if cataloguing rules would require an author/title access point.

If the field contained a record control number, then the program could proceed as follows when it produced the record in the catalogue from this record. If the record to which the link were made (that of the series) had a main entry under author, an author title added entry would be produced for this item in the series. If the record of the series on the other hand was entered under title, then a title added entry for the series would be produced in the record of the monograph as in the example below.

Record of monograph contains a link to a monographic series.

```
Label bibliographical level code: m
001    20055
010    $a92-2-106396-8
100    $a19890208d1988       f0ENGy0103a
101 0 $aeng
200 1 $aFrom a developing to a newly industrialised
country$ethe republic of Korea 1961-82$fTony Michell
210    $aGeneva$cILO$d1988
215    $axii, 180 p
225       2      $aEmployment,      adjustment      and
industrialisation$x0257-3415$v6
461  1$100120054$12001 $v6
700  1$aMichell$bTony

Record of series
Label bibliographical level code: s
001    20054
011    $a0257-3415
100    $a19890208s19869999   f0ENGy0103a
101 0 $aeng
200 1 $aEmployment, adjustment and industrialisation
210    $aGeneva$cILO$d1986
712 02$aILO$31092
```

Output in AACR form

Michell, Tony
From a developing to a newly industrialised country : the republic of Korea 1961 - 82 / Tony Michell. -- Geneva : ILO, 1988. -- xii, 180 p. -- (Employment, adjustment and industrialisation, ISSN 0257-3415 ; 6). -- ISBN 92-2-106396-8

**ADDED ENTRIES**
CORPORATE AUTHOR(S): ILO
SERIAL TITLE: Employment, adjustment and industrialisation
Record no: 20055

UNIMARC Authorities

From the outset, in many MARC formats, there had been problems of how to cope with references . LC MARC did not include them. UK MARC included in each record every reference required for all the headings in that record. The rationale behind that was that if you had taken only that one record with a particular heading, you would need to find all its references in that record to add them to the database. The logical way forward was for a format which would facilitate the setting up of databases of authority records. UNIMARC itself had incorporated in the access point fields a subfield, $3, which would allow the entry of a code which hopefully in the future would be an international authority number but for the present would be a number allocated to a heading in a particular system. It was not clear in the original manual or in the UNIMARC handbook how this would be done. Would the bibliographic records include the text of the headings and the codes, or would the headings be replaced by codes? The logical way to deal with access points in modern database management systems is to create separate records for each heading and link them to all the records in which they need to appear, calling them in to those records by means of the database number or some other identifier. However, this is not so convenient when exchanging bibliographic records since it is hard to ensure that all authority records are included in files along with bibliographic records. It is probably better to exchange records in complete form and include an authority number as well. If the records have originated from a source where an authority file has been used consistently, then the receiving system should be able to match them up, and perhaps replace them by authority records created from the names held as bibliographic data. So, the main reason for exchanging authority files is probably not to avoid having to include headings in bibliographic records, though it will obviate the need for bibliographic references to be included in the bibliographic record. Many organizations also wish to have access to authority files for their own record creation and the best way for them to obtain these from national agencies is in machine-readable form so that they may be used directly in their record creation and reduce the vast effort put into creating headings and their references.

To facilitate the exchange of authority information, in 1979 the IFLA Sections on Information Technology and Cataloguing jointly set up the IFLA Working Group on an International Authority System. This submitted in 1983 the Guidelines for authority and reference entries [6] (GARE) which set out the data elements that should appear in authority and reference entries in eye-readable form, using conventions akin to the punctuation in ISBD.

Then followed the development of a companion format, based on the underlying principles of UNIMARC and under the auspices of a Steering group on a UNIMARC Format for Authorities [7]. An additional principle was added, that subfield codes should be as in the bibliographic format, though the tags would have to differ because of the different functions of the fields in the different formats.

## CRITIQUE OF UNIMARC

Although UNIMARC has been adopted as a national format in many countries, it is intended as an international exchange format into which national agencies will convert their national records to cut down on the bilateral conversion arrangements in which national agencies would otherwise have to engage.

As an international exchange format, it had to be able to cater for all the idiosyncracies of existing national formats.

For this reason, the UNIMARC format contains some redundancy; one reason why the UNIMARC handbook was commissioned was to give users of UNIMARC guidance as to which option to take in those circumstances where data could be transferred from one field in a national format to two in UNIMARC. One can see a certain amount of overlap between Uniform Titles, Collective Uniform Titles, Uniform Conventional Headings and Topical Name Used as a Subject.

Because records created under different cataloguing rules may be held in the UNIMARC format, it is difficult to cater for every eventuality. Some cataloguing codes, increasingly as adaptations are made for automation, may not have the concept of main entry. So a way has to be included to code these records as UNIMARC does cater for main entry. Unfortunately, records using main entry and those that do not will never be completely compatible. But compatibility is a relative concept and it is well-known that if we want to share records we always have to make some compromises.

Another criticism that has been made of UNIMARC is that it is not promoted enough. Some of the orgnaizations instrumental in its establishment, for example the British Library, have not used it. For developing countries to use it means an expensive outlay for documentation. The working groups that have set it up have been composed of experts from organizations who have already been users of national MARC formats and there has not been anyone from developing countries for whom it is also intended since participants of IFLA working groups always have to pay their own expenses. IFLA has begun to think about this and in future they hope to subsidise representatives from developing countries and they also hope to hold UNIMARC workshops, some in developing countries. The first of these was held during the IFLA General Conference when it was held in Australia in 1988 [8], and a similar event is to take place in Cooperation with Unesco in Florence in late Spring 1991.

and it was made clear that this format should not be used for serials by excluding the category of 'serial only' from the table, and to exclude holdings data.

After publication, it was felt that the manual needed a maintenance agency to look after it and so the UK government, anxious to avoid being upstaged by the French government which had set up the ISDS Centre, agreed to host a UNISIST Centre which was named UNIBID. After hosting the Centre for over five years, the British Library lost interest in the project and transferred the functions of UNIBID to the Unesco Division of the General Information Programme which continued to provide copies of the manual to enquirers. However, the second edition which had been published in loose-leaf format was not updated as such because of shortage of staff and the labour intensive nature of the distribution of loose-leaf publications, and this edition was superseded by a third edition incorporating all the changes in 1985.

The manual was widely circulated by Unesco and it exerted extensive influence on systems that were being developed in the 1970's and early 1980's. It was used as a source of data elements by organizations developing formats. It was first used by CEPAL (UN Comision Economica para America Latina) in Latin America, where a format was developed with tags in a one-to-one relationship with those of the Reference Manual. But the system used only two-digit tags, as it was designed to work with ISIS on IBM mainframes. The CEPAL format is probably the most widely used in Latin America. The Reference Manual format was used by the International Development Research Centre in Ottawa as a format on which to model the format for DEVSIS, the Development Information System, and was then adopted for the MINISIS software system [12]. This package, developed by IDRC as a package to be made available to organizations in developing countries for their library databases is prominent among software packages in following the Reference Manual in having four-digit alphanumeric tags (one alphabetic character followed by three numeric, the last of which is a subfield identifier). The package has only recently had additional software written for it to enable it to support ISO 2709-based formats which have the usual three-digit tags. Users of the package were encouraged to use their own fields and field definitions, since it was part of the philosophy of IDRC that nothing should be imposed on users from above, though reference was made in the manuals to documents like the Reference Manual and the use of official international standards has always been encouraged.

A further interesting success story involving the Manual is that of the American Geological Institute's abstracting service GeoRef. This organization was one of the first agencies to adopt the Reference Manual as the basic format of its automated bibliographic information system. They specialize in indexing all English Language material in their subject field. Mulvihill tells [13] how when they decided to extend the coverage to French material by means of a co-operative agreement with CNRS in France, they had no difficulty in merging files with each other; since CNRS had been heavily involved in the design of the Reference Manual, its format was compatible with that of GeoRef.

## Technical features of the format

The major feature of the format is that it gives equal prominence to bibliographic records whether they relate to analytics (meaning journal articles and contributions in journals as well as works found published separately elsewhere but here bound together), monographs or serial titles. The format was designed to do this because it was developed by secondary services which give equal prominence to the different bibliographic levels. It does this in a so-called 'flat' record structure. The record contains no distinctive feature to permit a hierarchy to be indicated; instead, different tags are allocated to fields at a particular level. Thus, a computer program interpreting the record has to hold a table in which each field is separately identified. Additionally, certain fields such as ISBN and publisher are not identified as belonging to any particular bibliographic level; in most cases the level of these fields is implied, as publisher, for example, relates to the monograph. As mentioned above, the group developing the format avoided enabling the format to be used for serial titles, and in the matrix in the first edition giving combinations of fields for types of material there is no column for serial title. Tag A08 is the field identifier for title of analytic, A09 title of monograph and A10 title of collection level. A03 is the field for title of serial. In the second edition of the Reference Manual, the scope of fields A13 and A19, (Person and corporate body associated with collection) has been extended to include responsibility for serials.

# UNESCO COMMON COMMUNICATION FORMAT

## History

Although Unesco had developed the Reference Manual with the help of ICSU/AB, it had not been accepted unquestionably by the audience it was intended to serve. Many organizations continued to approach Unesco for assistance in developing bibliographic information systems; sometimes these organizations were related to national libraries and needed to establish data bases that were compatible with MARC. Sometimes they were organizations that straddled the divide conventionally believed to exist between the libraries and secondary services. Some were even situated within national libraries but were secondary services, so it was difficult to see whether they should follow the Reference Manual developed for the secondary services or UNIMARC, developed by and for national libraries. In order to solicit wider opinion on the problem and thereby to help in its decision making, Unesco sponsored the International Symposium on Bibliographic Exchange Formats. This took place in Taormina in April 1978 and was organized by UNIBID, the office supported by the Unesco General Information Programme and the British Library which was responsible for maintaining the Reference Manual. The Symposium also enjoyed the sponsorship of ICSU/AB, IFLA and ISO. Papers were given on a number of issues relating to the then state of the art of exchange formats and outlines were given of the main features of the major international formats. The proceedings were published in late 1978 [14]. As a result of resolutions passed at the

Symposium, Unesco set up the Ad hoc Group for the Establishment of the Common Communication Format. This Group contained experts from ICSU/AB, ISDS (the International Serials Data System), IFLA, ISO and UNIBID, as well as an expert from the group that had devised MEKOF, the format of the CMEA (Eastern European) countries [15]. The Group worked on the basis that the new format must be compatible with the MEKOF, UNIMARC and UNISIST Reference Manual formats. It also took into account derivatives of these formats, namely the USSR/US Exchange Format (based on UNIMARC) and an ICSU/AB Extension to the Reference Manual developed by the Four Ways Committee. The Group agreed that the record structure of the format should be that specified in the ISO 2709 standard, which was in any case used by all the formats being taken into account. A consultant prepared a data element directory which included the majority of the data elements from those formats.

In the early days of the Group, much of the discussion centred on the adoption of a basic set of mandatory data elements. It was clear that the secondary services were not prepared to adopt the mandatory elements of ISBD. For instance, the statement of responsibility was not provided by many of their databases. The libraries community was persuaded that, though the ISBD elements were, in principle, desirable, records without certain of the elements from sources without the tradition of fullness of the record that is found in the national libraries would nevertheless be useful to such libraries. The format was aimed at operations which needed to provide records to and receive records from both library and secondary service community, and as many of these organizations were in developing countries, it was decided to keep the format simple in terms of its data elements and data element definition. Taking into account the fact that there was not then, and indeed still is not, any international agreement on cataloguing rules, the format was kept free of anything amounting to cataloguing rules. In order to achieve compatibility between the different record structures of the formats and their differently-defined bibliographic levels, a record structure was defined for the CCF implementing the latest version of ISO 2709. The structure of the format has at times been criticized as over-complex. It might be true that it is not easy for cataloguers to understand: that is because it requires a different approach from that of traditional cataloguing on which, incidentally, secondary services practices also are usually based. However, the CCF is, as a standard, only required to be implemented as an exchange format, so the total computerized system should take this into account, and allow records to be created in a way that more closely resembles data entry practices in other automated systems. This will require a data entry format which is different from the exchange format. It may be obvious to many users that this can be done to simplify data entry. However, there are other users who are still of the opinion that to follow the CCF it is necessary to use the data elements as described in the manual, and their identifiers, at every possible level in the system. This is possible for the MARC formats as they were developed to automate existing manual systems geared up to the production of catalogue cards. The CCF on the other hand was designed from a data element directory.

The format was published in 1984 [16].

UNISIST REFERENCE MANUAL

History and use

UNIMARC and the MARC formats have been developed for the library sector of the information community.

Computers were already being employed by secondary services before they were introduced into libraries. In the context of the exchange of data the secondary services were to follow the libraries. Since the record structure of the MARC format had been made an international standard ISO 2709 [9], it was the obvious standard for the information community as a whole to follow. In the United States, the Chemical Abstracts Service followed the Library of Congress in setting up a similar cooperative project to that which the Library of Congress had set up with the British National Bibliography, this time with UKCIS, the UK Chemical Information Service. They, too, took the MARC record structure as the standard record structure. In the UK, the Institution of Electrical Engineers started in 1969 a tape service for bibliographic references, automating their abstracting and indexing service which began as Science Abstracts in 1898. This, too, used the same record structure even before any thought had been given to adopting it as a standard. The need for a standard set of data elements for the exchange of bibliographic data was spreading to the secondary services, so they began to look for something akin to the MARC formats. They based their format on the same record structure, though they adopted their own system of tags for the data elements.

Resolutions adopted at the 14th and 15th Sessions of the General Conference of Unesco which took place in 1966 and 1968 authorized the Director-General of Unesco to undertake and complete jointly with the International Council of Scientific Unions (ICSU) a feasibility study on the establishment of a World Science Information System (UNISIST) [10].

The UNISIST-ICSU/AB Working Group on Bibliographic Descriptions, set up in 1967 as part of the UNISIST programme decided that it was necessary to develop a standard for the recording and exchange of data in machine-readable form. The outcome of this was the UNISIST Reference Manual for Machine-Readable Bibliographic Descriptions [11] and the group that had worked on it included representatives from the the British National Bibliography, the Centre National de Recherche Scientifique, France, the Institution of Electrical Engineers who had set up INSPEC, and Chemical Abstracts.

When the format was being developed, the Working Group had only the early MARC formats as models. The members decided that they should take great care not to cause confusion with the existing MARC formats and decided that tags should begin with an alphabetic character, and subfield identifiers should be numeric. Because the International Centre of the International Serials Data System was engaged in the control of serial titles, it was decided that the Reference Manual should not include the treatment of serials as a whole, so no provision was made for them. However, fields were included for the treatment of contributions in serials. The Manual included matrices or tables giving the fields required for each combination of bibliographic level (e.g. analytic in monograph in series; monograph; monograph in series)

Users of the Format

Even before the format was formally published, two major organizations were already using it. The Dag Hammarskjold Library of the UN in New York adopted the CCF. A data entry manual has been published, the UNBIS Reference Manual [12].

The Office of Official Publications of the European Communities was developing new software and adopted the CCF because of its flexible record structure. They were interested not only in providing a mechanism for linking bibliographic records to each other but also in providing the facility for the linking of the actual text. They publish the Official Journal of the European Communities which consists of small items of information in a daily journal with weekly supplements. These have been put in a large database, each item including its text constituting one record. The main aim is to enable the journal to be printed from tapes in different centres throughout the European Community. The bibliographic levels and segments of the CCF have been used to the full to enable the data from the different sections in the publication to be arranged in their appropriate segments. FORMEX has been published and from the document it can be seen that it adheres very closely to the CCF.[18]

Probably the first network to adopt the CCF was the ICONDA Group developing an international construction database. They had originally planned to use the UNISIST Reference Manual, but, because they were intending to merge databases which had already adopted data entry rules, they found the CCF easier to implement and have based their manual on it [19].

Since publication of the CCF, a number of organizations have been helped by Unesco to investigate the advantage of using the format, and, where it has proved advantageous, to adopt it in one way or another.

Simmons [20] relates how in Colombia COLCIENCIAS, a semi-autonomous government agency took on the task of creating and co-ordinating a co-operative national information system to include the resources of documentation centres, libraries and archives, many of which were microcomputer based. These organizations were separately funded and chose their own computer hardware and software. A 'switching format' based on the CCF has been designed called the Formato Comun de Comunicacion Bibliografica para Colombia (FCCC). Each participating agency required a pair of programs to be written, to convert its records to FCCC and back. Programs will also enable the conversion from FCCC to CCF and back. In Venezuela, there is a desire to follow this pattern since there are many users of the MARC and CEPAL formats. Howveer, those who wish to set up databases on Micro-ISIS prefer the CCF as they find it has just the right level of flexibility for their needs [21].

The International Co-ordinating Committee for Development Associations (ICCDA) has developed an implementation of the CCF on the CDS/ISIS Microcomputer Software Package which is intended for producing databases which can be exchanged between participants. A manual accompanies the software package [22]. The work on the package was co-ordinated by the OECD Development Centre and supported by IDRC. This package is being used as a model for other similar implementations outside the development community wishing to use the CCF and the CDS/ISIS package.

In China, too, the CCF has been translated and is beginning to be promoted in organizations that need to participate in both the library and secondary service the library and the secondary services community and in a Chinese translation was begun in 1989.

The second edition of the format was published in May 1988 [23], and in April 1989, the first Users Meeting took place at the International Bureau of Education in Geneva, sponsored by Unesco, at which progress reports, technical papers and practical demonstrations were given on topics such as implementing the CCF on particular software systems, future extensions to the format for additional kinds of material and conversions between the CCF and other formats [24]. The next edition of the format will probably be published in 1991 or 1992 and will include a twin manual for factual data, initially research projects, persons and instiutions. The CCF (Bibliographic) will most probably also be revised and will include fields for cartographic materials, standards and patents. Close liaison is taking place between the working group and the UNIMARC community to ensure that the CCF remains compatible. An integrated database on the software package CDS/ISIS for Microcomputers, including the facility to hold bibliographic as well as factual data is under development and it will include a user manual. It is likely that this will be circulated with CDS/ISIS when it becomes available as an additional standard database to the database supplied with the package at present.

## Technical aspects

As mentioned above, the record structure of the CCF has been criticized as over-complex. In fact, as a machine-readable format it is the opposite, and it can be thought of as complex only when it is regarded as a data entry format which it was not intended to be. It is complicated for cataloguers to enter data into the format, especially if they try and create manually the links between records or between segments in a record.

There are two main features of the format that distinguish it from other formats. The first feature is its simple set of data elements that can be used at any bibliographic level and are disassociated from cataloguing codes. The second is the logically-defined record structure which uses the fourth element of the ISO 2709 directory to denote bibliographic level and field occurrence. The use of both of these features is a product of the circumstances in which the format was devised. Since the format was designed to be compatible with a number of other already existing international formats, it was necessary either to include all data elements from these other formats, or a subset. Including all data elements, in particular those that are seldom used, would have decreased the level of compatibility in the CCF. It is in the lesser used data elements that the formats have gone their own way. Therefore it was decided to include the basic elements in the format for exchange and let the less commonly used data elements be added as private data elements between parties to an exchange agreement. Another reason for there being fewer data elements than there would otherwise be is that data elements relating to different bibliographic levels are not allocated to different fields at each level but appear only once as one field. Field 200 is the field for title. If the title is the title of a monograph,

it will be designated to a segment containing all the fields relating to the monographic level. If the title is that of an article it will be designated to a segment containing all the fields relating to that article.

The record structure of the CCF was devised to take into account different structures in the format from which records would originate. The Reference Manual and formats related to it have fields designated for different bibliographical levels. UNIMARC has fields designed primarily for the monographic and serial level but can also use those fields embedded in linking fields as fields describing an analytic. The Reference Manual has four bibliographic levels, analytic, monograph, serial and collective, whilst UNIMARC has analytic, monograph, serial and collection. Collective in RM corresponds to multi-volume monograph in UNIMARC (only a subset of monograph). In both source formats, the fields relating to appropriate bibliographic levels can easily be identified. However, the relationships could more easily be converted into a third more logical structure than into the structure of the other of the original formats, so the structure of the CCF was designed to be logical. It was designed to make use of a then new feature of ISO 2709, the fourth element of the record directory, so that each field is denoted (in this fourth part of the directory) as belonging to its bibliographic level and each field in the record is uniquely identified there by an occurrence identifier.

Field to field links have also been included in the CCF. The second edition includes codes to denote links between an author name and his affiliation (which will usually be entered in its own field and may be formatted like a corporate body if the rules permit) and between publisher and ISBN where a record includes two publishers of a simultaneously published work.

The next edition of the CCF will include a new type of link, record-to-record link, which will obviate the need to use segments when links are being made from one record to another.

In evaluating the CCF it is necessary to remember three points:

a) Relationship with existing formats

The CCF was not designed from first principles but was based on major existing international exchange formats and was intended to be used for the transfer of records between systems which were already capable of providing output into the these major exchange formats.

It was not expected to have to do anything that could not be done by any existing exchange format.

It is possible to take a bibliographic item such as a series of annual conference proceedings where each member of the series has its own individual articles and create one record containing all the data relating to what would amount in most bibliographic systems to a number of records. However to comply with the CCF, this record will contain a segment for each separately occurring instance of each bibliographic level. One of these segments has to be labelled the primary segment and this will contain certain elements of control information such as record control number. If the format had been designed from first principles it would have probably contained a control segment in each record which

would always be present and would contain information as to which segments would make up a complete bibliographic record. As it is, it is the primary segment which contains this control information.

b) The CCF is an exchange format

The CCF is intended as an exchange format and as such has to contain bibliographic data for exchanging between systems. It does not govern what can be done within the systems themselves, so it cannot be looked to as a guide for creators of on-line public access catalogues or other systems. Of course, the definition of data elements will affect the internal architecture of systems using these data elements, but there is a large amount of agreement between organizations as to the definition of the key data elements in a record. This can be noted by comparing the data elements in a national bibliography and in a secondary service publication. The data elements author, title, publisher, date, to mention only a few, will be there in every case although they may be presented in different forms, according to different cataloguing codes.

c) The CCF is intended for exchange of bibliographic data

Thirdly, when the system was developed it was intended for the exchange of those data elements of the bibliographic record that were needed for the identification of a document in a catalogue or bibliography. It does not contain fields that would be required for library circulation systems or inter-library loan. An individual system using the CCF as an exchange format to facilitate record creation by taking records created externally in the CCF may add any other fields required for its own purposes. Moreover, systems wishing to exchange data elements other than those provided for in the CCF are free to allocate unused tags to those data elements or to allocate alpha-numeric tags (e.g. AAA, BAZ, H97).

## CONCLUSION

There is little to say in conclusion. Only by a study of the different exchange formats and an investigation of the users can a decision be made as to the format on which one's own system should be based. If you have to exchange data with organizations in both the library and secondary service community then most probably the CCF is for you. If you are a secondary service and want to give equal treatment to articles as to books and reports then the UNISIST Reference Manual will serve your purpose. If you are an academic library and want to exchange data with your national library then you should probably adopt the format used by the national library. The chances are that that will be based on UNIMARC or US or UK MARC, though, again, some national libraries that straddle the divide between libraries and secondary services especially in developing countries have adopted the CCF.

References


1 ISO 2709 : Documentation : format for bibliographic information interchange on magnetic tape. 2nd ed. Geneva, ISO, 1981
2 UNIMARC. 2nd ed. London, IFLA Office for UBC, 1981. p1
3 UNIMARC handbook. London, IFLA Office for UBC, 1983
4 UNIMARC manual. London, IFLA Office for UBC, 1987
5 IFLA journal 13 (1987) p 168
6 Guidelines for authority and reference entries / recommended by the Working Group on an International Authority System. London, IFLA International Programme for UBC, 1984
7 UNIMARC / A -- UNIMARC format for authorities. London, IFLA UBCIM Programme, 1990.
8 UNIMARC in theory and practice: papers from the UNIMARC workshop. London, IFLA UBCIM Programme, 1989
9 ISO 2709. op.cit.
10 ISDS Manual. Paris, ISDS International Centre, 1983. p1
11 Martin, M.D. (ed) Reference manual for machine-readable bibliographic descriptions. Paris, Unesco, 1974 (SC.74/WS/20) Dierickx, H & Hopkinson, A. Reference manual for machine-readable bibliographic descriptions. 2nd rev. ed. Paris, Unesco, 1981
12 Woolston, John. DEVSIS: a development science information system. Ottawa, IDRC, 1974
13 Mulvihill, J.G. GeoRef coverage and improvements in the bibliography and index of geology. In Prewett, N.J. (ed.) Keeping current with geoscience information. Washington, DC, 1981. pp55-64
14 International Symposium on Bibliographic Exchange Formats. Towards a common bibliographic exchange format?: proceedings. Budapest, OMKDK, 1978
15 International interchange format: MEKOF-2: specification of data elements. Moscow, ICSTI, 1977 (MEKOF-2/SPEC (Ed.1) 1977)
16 CCF: the Common Communication Format. Paris, Unesco, 1984
17 UNBIS: Reference manual for bibliographic description: a manual for the preparation of bibliographic data for input into and retrieval from the United Nations Bibliographic Information System. New York, UN Dag Hammarskjold Library, 1985
18 Guittet, C. (ed.) FORMEX: formalized exchange of electronic publications. Luxemburg, Office for Official Publications of the European Communities, 1985 ISBN 92-825-5399-X
19 ICONDA communication format: format for the exchange of records in the frame of the International Construction Database. Stuttgart, IRB Verlag, 1985
20 Simmons, P. Using CCF: the Common Communication Format, in Information technology and libraries 5(4) 1988 p285-294
21 Hopkinson, A. A common bibliographic exchange format for Venezuela. Brighton, IDS, 1990
22 Di Lauro, Anne. IDIN manual for the creation and management of a bibliographic database using Micro-ISIS. Paris, OECD, 1988. 189p includes diskette
23 CCF: the Common Communication Format. 2nd ed. Paris, Unesco, 1988
24 Simmons, P.(ed.) Proceedings of the First CCF Users' Meeting. Paris, Unesco, 1990